

Data-Based Fuzzy TOPSIS for Alternative Ranking

Victor Utomo^{a*}, Rachmat Gernowo^b

Information Technology Department of STMIK PROVISI, Semarang,
Science and Mathematic Faculty of Diponegoro University, Semarang

Naskah Diterima : 30 Desember 2012; Diterima Publikasi : 12 April 2013

Abstract

Technique for order preference by similarity (TOPSIS) solves multi-criteria decision making (MCDM) by ranking the alternatives. When the attributes are not deterministic, a Fuzzy TOPSIS method is applied. The traditional fuzzy TOPSIS depends on decision makers to determine alternative's value which considered subjective. A new method named data-based fuzzy TOPSIS proposed to diminish the dependency to decision maker. The proposed algorithm use data to determine alternative's values objectively. Subtractive Clustering (SC) and Fuzzy C-Mean (FCM) selected to transform crisp value data to fuzzy value data. Some modification applied to SC and FCM to obtain fuzzy triangular value needed by fuzzy TOPSIS.

Keywords: Decision support systems; Fuzzy TOPSIS; Fuzzy C-mean; Subtractive clustering

Abstrak

Technique for order preference by similarity (TOPSIS) memecahkan masalah multi –criteria decision makin (MCDM) dengan memberikan peringkat untuk setiap alternatif. Metode Fuzzy TOPSIS digunakan jika atribut yang terlibat bersifat non deterministik. Fuzzy TOPSIS pada umumnya bergantung pada penilaian pengambil keputusan secara subyektif untuk menentukan nilai alternatif. Pada penelitian ini diusulkan metode Data-Based Fuzzy TOPSIS untuk menghilangkan ketergantungan pada pengambil keputusan. Algoritma yang disulkan menggunakan data untuk menentukan nilai alternatif secara obyektif. Transformasi dari nilai crisp ke nilai fuzzy menggunakan method Subtractive Clustering (SC) dan Fuzzy C-Mean (FCM). Perubahan dilakukan pada metode SC dan FCM untuk mendapatkan bilangan fuzzy segitiga yang digunakan pada Fuzzy TOPSIS

Kata kunci : Sistem pendukung keputusan; Fuzzy TOPSIS; Fuzzy C-Mean

1. Introduction

Certain problems are considered multi-criteria decision making (MCDM) when decision not determined from single point of view. Technique for Order Preference by Similarity (TOPSIS) solves MCDM the problem by ranking the alternatives. As the alternative getting closer to the positive ideal solution (PIS) and further from the negative ideal solution (NIS), the alternative's rank gets higher. In problems when the attributes are not deterministic, the fuzzy logic is usually adopted. Both circumstances create fuzzy TOPSIS (Madi *et al.*, 2011)

While fuzzy TOPSIS is proven to solve problems in MCDM area (Madi *et al.*, 2011), (Alborxi, 2008) there is dependency to decision makers. Decision maker subjectively determine the value of decision matrix which affects the result. To improve objectivity, the value of decision matrix is usually determined by more than one decision makers. The traditional algorithm also need decision maker to determine the value of decision matrix when new

alternative added to the problem.

Overcoming the dependency to decision maker, the research proposed a new algorithm. Instead from decision maker, the proposed algorithm obtains input from data. While fuzzy part of the traditional algorithm is kept since considered appropriate to the nature of non deterministic problem, transformation from crisp value to fuzzy value is needed. Combination of subtractive clustering and fuzzy C-mean is proposed for the transformation.

The rest of the paper structured as follows, section 2 describes the methodologies used in the research and section 3 explains the proposed method. An example is taken and the result discussed in section 4. In section 5 the research is concluded and some suggestions are given.

2. Research Methodologies

2.1. Subtractive Clustering

Subtractive clustering (SC) is a fast, one-pass algorithm for estimating the number and the centers of clusters for a set of data (Hossen, 2011).

*) Penulis korespondensi: victor.utomo@ gmail.com

Given a set of data, $D = \{z_d | d = 1, 2, \dots, n\}$, set the normal value based on z_{d-n} and z_{d-m} use the following model (Priyono, 2005).

$$z_d^{norm} = \frac{z_d - z_{d-min}}{z_{d-max} - z_{d-n}} \quad (1)$$

Process the normalized data use the following algorithm

Step 1. Calculate possibilities of each data to become a cluster center use the following formula

$$P_d = \sum_{j=1}^n e^{-\alpha \|z_d - z_j\|} \quad (2)$$

where $\alpha = \frac{1}{r_d}$ is a positive constant, and $\| \cdot \|$ is the Euclidean norm. The constant r_d is effectively the radius defining a neighborhood, a data point with many neighboring data points will have a high potential value and therefore is likely to be selected as the cluster center.

Step 2. Let $K = 1$. Find the first cluster center

$$z_1^* = \{z_j | P_j = \max_d P\} \quad (3)$$

Step 3. Update

$$P_d = P_d - P_d^* e^{-\beta \|z_d - z_1^*\|} \quad (4)$$

where $\beta = \frac{1}{r_d}$ is a positive constant. The constant r_d is effectively the radius defining the neighborhood within which data points will have significant reductions in potential and therefore are unlikely to be selected as the next cluster center.

Step 4. $K = K + 1$. Choose the K th cluster center candidate

$$z_K^* = \{z_j | P_j = \max_d P\} \quad (5)$$

Step 5.

If $P_K^* > \varepsilon_h$

Accept z_K^* as a cluster center and go to step 3.

Else if $P_K^* \leq \varepsilon_l$

Reject z_K^* and end the clustering process.

Else

Let d_{min} shortest of the distance between z_K^* and all previously found cluster centers.

If $\frac{d_{min}}{r_K} + \frac{P_K^*}{P_1^*} \geq 1$

Accept z_K^* as a cluster center and go to step 3.

Else

Reject z_K^* and set $P_K^* = 0$.

$K = K - 1$ and go to step 4.

End if.

End if.

Here ε_h specifies a threshold for the potential above which we will definitely accept the data point as a cluster center; ε_l specifies a threshold for the potential below which the data point is rejected (Hung, 1999).

Four parameters, α , β , r_d , and ε_h influence the number of clusters and accuracy of clustering. Usually the values for each parameter are

$$\alpha = 1.5, \quad 0.15 \leq r_d \leq 1, \quad \varepsilon_l = 0, \quad \text{and} \quad \varepsilon_h = 1 \quad (7)$$

2.2. Fuzzy C-Mean

Fuzzy C-mean (FCM) is an unsupervised classification method, belonging to the partition clustering category. The algorithm based on the minimization of the sum of squared Euclidean distance between data and cluster centers, using the following objective function.

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - v_j\|^2 \quad (6)$$

Consider clustering the dataset $X = \{x_1, x_2, \dots, x_n\}$ where n is the number of data points. $V = \{v_1, v_2, \dots, v_c\}$ is the set of corresponding cluster centers in the dataset, where $c \in [2, n]$ is the number of clusters. $U = (\mu_{ij})_n \times c$ is a fuzzy partition matrix. μ_{ij} is the membership degree of data x_i to the cluster center v_j . Meanwhile, μ_{ij} has to satisfy the following condition

$$\mu_{ij} \in [0, 1] \quad \forall i = 1, \dots, n, \forall j = 1, \dots, c \quad (7)$$

$$\sum_{j=1}^c \mu_{ij} = 1 \quad \forall i = 1, \dots, n \quad (8)$$

$\|x_i - v_j\|$ denotes the Euclidean distance between x_i and v_j . Fuzziness index $m \in [1, \infty)$ used to control the fuzziness of membership of each datum (Yang, 2007).

The original algorithm leaves some things to consider. There is no rule to determine the number of clusters while it is essential to determine cluster validity. FCM tries to classify the data into the given number of clusters. In the proposed algorithm, the value of m is determined by the result of subtractive clustering.

The original algorithm suggests initialization of membership matrix U uses a random value while the performance of the FCM clustering is very dependent on the choice of the initial cluster center (Hossen *et al.*, 2011). To achieve a minimum number of iteration, initialization is made by updating membership matrix with the original algorithm against temporary cluster centers. The temporary cluster centers are the result of subtractive clustering.

The proposed algorithm need fuzzy triangular as explain further in section 3. Since the original algorithm is not explicitly defined the shape, some modifications according to 7-8 and are taken:

1. Set fuzziness index $m = 2$.
2. Set the lowest data $x_{min} = \min_i x_i$ to be the left most cluster center and set the highest data $x_{max} = \max_i x_i$ to be the right most cluster center.

3. Further process the original algorithm result

A create a set of triangular number corresponds to the cluster centers such

$$T = \{t_1, t_2, \dots, t_{n-1}, t_n\} = \{(v_1, v_1, v_2), (v_1, v_2, v_2), \dots, (v_{n-2}, v_{n-1}, v_n), (v_{n-1}, v_n, v_n)\} \quad (9)$$

b. Data x_i considered triangular number when the corresponding cluster center v_j is the nearest cluster center to x_i according to highest membership

degree $\max_j \mu_{ij}$.

$$x_i = \{t_k | v_k = \max_j \mu_{ij}\} \quad (10)$$

Procedure in the modified FCM has following steps :

1. Set the fuzziness index, read in the data set and find the maximum and minimum values.
2. Based on the result from the subtractive clustering algorithm, i.e. possible cluster centers, execute the following steps
 - a. Set the number of clusters
 - b. Consider possible cluster centers as temporary cluster center and update membership matrix with

$$\mu_{ij} = \frac{1}{\sum_{k=1}^m (d_{ij}/d_{ik})^{\frac{2}{m-1}}} \quad (11)$$

where $d_{ij} = \|x_i - v_j\|, i = 1, \dots, n$ and $j = 1, \dots, m$.

3. Calculate the fuzzy centers using

$$v_j = \frac{\sum_{i=1}^n (\mu_{ij})^m x_i}{\sum_{i=1}^n (\mu_{ij})^m}, \forall j = 1, \dots, m \quad (12)$$

but force the two clusters with the largest and smallest values to take the maximum and minimum domain values.

4. Update the membership matrix with

$$\mu_{ij} = \frac{1}{\sum_{k=1}^m (d_{ij}/d_{ik})^{\frac{2}{m-1}}} \quad (13)$$

TABLE I
WEIGHT TO TRIANGULAR NUMBER VALUE

where $d_{ij} = \|x_i - v_j\|, i = 1, \dots, n$ and $j = 1, \dots, m$.

5. Repeat step (3) and (4) until one of the termination criteria is satisfied. Termination criteria either that the difference between updated and previous is less than a predefined minimum threshold or the iteration maximum number reached.
6. Process membership matrix using equation (9) and (10) to get the desired result.

2.3. Fuzzy TOPSIS

Let MCDM problem has alternatives A_1, A_2, \dots, A_n and criterias C_1, C_2, \dots, C_m . Each alternative will take a consideration with respect to criteria m. The ratings of criteria can be concisely expressed in matrix format as $X = (x_{ij})_{n \times m}$ and

$W = (w_1, w_2, \dots, w_m)$ where $x_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ and $w_j (j = 1, 2, \dots, m)$ are the fuzzy rating of alternative $A_i (i = 1, 2, \dots, n)$.

1. Decision matrix, $X = (x_{ij})_{n \times m}$ is normalized via equation (14)

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^m x_{kj}^2}}, i = 1, \dots, n; j = 1, \dots, m \quad (14)$$

This step is considered optional, only if the data are not in normal form, i.e. ranging from 0 to 1.

2. Positive ideal solution (PIS) and negative ideal

solution (NIS) are determined

$$A^* = \{v_1^*, \dots, v_m^*\}$$

$$= \{(max_j v_{ij} | j \in \varphi_b), (min_j v_{ij} | j \in \varphi_i)\} \quad (15)$$

$$A^- = \{v_1^-, \dots, v_m^-\}$$

$$= \{(min_j v_{ij} | j \in \varphi_b), (max_j v_{ij} | j \in \varphi_i)\} \quad (16)$$

3. The distance of each alternative from PIS and NIS are calculated using Euclidean distance formula

$$D_i^+ = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^*)^2}, i = 1, \dots, n \quad (17)$$

$$D_i^- = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2}, i = 1, \dots, n$$

Table 1. Weight to triangular number value

Weight	Value
Unimportant (U)	(0, 0.25, 0.5)
Neutral (N)	(0.25, 0.5, 0.75)
Important (I)	(0.5, 0.75, 1)

(18)

4. The closeness coefficient of each alternative is calculated

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-}, i = 1, 2, \dots, n$$

(19)

5. By comparing CC values, the ranking of alternatives are determined. The nearer the value close to 1 implies alternative approach to positive ideal solution, i.e. the maximum value of, then the optimal alternative can be ranked by decision maker. Finally, the best alternative can be selected (Azadnia *et al.*, 2011).

Table 2. Dataset in experiment

Alternative	And Ver	Weight	Proc Speed	Price
Sony Ion LTE LT28i	2.3	144	1500	5600
Sony Xperia Acro S LT26W	4	147	1500	5100
Sony Xperia Go ST27i	2.3	110	1000	2975
Sony Xperia J ST26i	4	124	1000	2650
Sony Xperia Miro ST23i	4	110	800	2100
Sony Xperia Neo L MT25i	4	131.5	1000	2650
Sony Xperia P LT22i	4	120	1000	4150
Sony Xperia S Nazomi LT26i	2.3	144	1500	4900
Sony Xperia SL	4	144	1700	5100
Sony Xperia Tipo ST21i	4	99.4	800	1600
Sony Xperia U Kumquat ST25i	2.3	110	1000	2299

2.4. Proposed Algorithm

Data-based fuzzy TOPSIS for alternatives ranking consist of two sub-systems, converts crisp data to fuzzy form and creates alternative ranking with fuzzy TOPSIS, as shown in fig. 1. System's input is data in crisp value. The crisp value transform into fuzzy number. Since the proposed algorithm use triangular fuzzy number, the result of fuzzification process must meet the requirement. The fuzzy data processed by fuzzy TOPSIS to obtain ranked

alternatives.

Some parameters, i.e. criteria, weight, PIS, NIS and r_a , must be set to use the system. Compare to the original FCM, a new parameter is introduced. Parameter criteria, weight, PIS and NIS need expert to supply the value, while parameter α determined in a more technical manner.

The proposed algorithm to build data-based fuzzy TOPSIS is described as follow:

1. Choose a set of criteria, the corresponding weight, positive ideal solution and negative ideal solution to each criterion and α .
2. Retrieve data from database. Put the data in the decision matrix in the following manner, data fields and rows correspond respectively to system criteria and alternatives.
3. Repeat the following steps for each criterion
 - a. Get a set of possible cluster center use SC algorithm described in section II.A.
 - b. Use the result of SC algorithm and crisp data as input of FCM algorithm described in section II.B to have the data in the form of fuzzy triangular number.
4. Creates alternatives ranking with fuzzy TOPSIS method use the result from step 3.

3. Example and Discussion

An experiment in ranking of Sony Mobile Android Phone is taken as an example. Triangular number value for each weight is defined in table 1.

In the experiment, set parameter $\alpha = 0$, while criteria parameters are described in table 2.

Two datasets were taken to inspect system behavior to different dataset. The first dataset shown is the first 8 row shown in table III while the second dataset is all the data shown in table 3.

Result set from first dataset and second dataset are shown respectively in table 4 and table 5.

Table 4. First result set

Rank	Alternative	CC
1	Sony Experia J ST26i	0.6703
2	Sony Experia Miro ST23i	0.6499
3	Sony Experia Neo L MT25i	0.6401
4	Sony Experia Go ST27i	0.5694
5	Sony Experia P LT22i	0.5155
6	Sony Experia Acro S LT26W	0.4646
7	Sony Experia S Nazomi LT26i	0.3777
8	Sony Ion LTE LT28i	0.3501

Table 5. Second result set

Rank	Alternative	CC
1	Sony Experia Tipo ST21i	0.6399
2	Sony Experia Miro ST23i	0.5997
3	Sony Experia U Kumquat ST25i	0.5825
4	Sony Experia J ST26i	0.5485

5	Sony Experia Neo L MT25i	0.5301
6	Sony Experia SL	0.4760
7	Sony Experia Go ST27i	0.4618
8	Sony Experia Acro S LT26W	0.4519
9	Sony Experia P LT22i	0.3981
10	Sony Experia S Nazomi LT26i	0.3619
11	Sony Ion LTE LT28i	0.3238

Set parameter $\alpha = 1$ on the first dataset gives third result set shown in table 6.

Table 6. Third result set

Rank	Alternative	CC
1	Sony Experia J ST26i	0.7384
2	Sony Experia Neo L MT25i	0.6982
3	Sony Experia Miro ST23i	0.6208
4	Sony Experia Go ST27i	0.6157
5	Sony Experia P LT22i	0.5444
6	Sony Experia Acro S LT26W	0.4715
7	Sony Ion LTE LT28i	0.3792
7	Sony Experia S Nazomi LT26i	0.3792

Comparison of first and second result set shows that the proposed algorithm change the result under data modification without intervention of decision maker while in the traditional fuzzy TOPSIS decision maker intervention needed to change the result. The fact confirms the ability of the proposed algorithm to behave more objective compared to the traditional fuzzy TOPSIS.

The first and third result sets show the effect of α value difference. While the optimal value of α is out of research's scope, experiment shows that a lower value of α gives a better resolution of closeness coefficient.

Since the proposed algorithm depends on the data, some consideration about data is made. Averaging last couple month data gives different result than averaging last year data. Under the circumstances, a different case may need different query. It means that data query optimization is required. Another consideration is about the data source, preprocessing transactional database may be required to clean the data.

4. Conclusion

Data-based fuzzy TOPSIS method works in a more objective manner compared to the traditional fuzzy TOPSIS method. Change in data affects the result without decision maker intervention. The proposed algorithm works by combining SC, FCM and fuzzy TOPSIS. SC and FCM used to create a dynamic fuzzification of the data.

The proposed algorithm depends on the parameters, criteria, weight, PIS, NIS and Like the traditional fuzzy TOPSIS, parameters criteria, weight, PIS and NIS are determined subjectively by experts. To achieve a more objective result, another study about the parameters is required. Optimization

of parameters and data query also required to achieve a better result.

Experiments show that at certain situation the proposed algorithm fails. Further examination shows that when the data are not well distributed, clustering fails. Work in the area needed to overcome the issue.

References

- Madi, E.N. and Md. A.O., 2011. Fuzzy TOPSIS method in the selection of investment boards by incorporating operational risks, in *2011 Proc. World Congress on Engineering.*, pp. 291-295.
- Uyun, S. and Riadi, I., 2011. A fuzzy multiple-attribute decision making for scholarship selection, *TELKOMNIKA*, 9(1): 37-46.
- Ding, J.F., 2011. An integrated fuzzy TOPSIS method for ranking alternatives and its application, *Journal of Marine Science and Technology*, 19(4): 341-352.
- Salehi, M. and Tavakkoli-Moghaddam, R., 2008. Project selection by using a fuzzy TOPSIS technique, *Int. Journal of Computer and Information Engineering*, 2(2): 100-105.
- Azadnia, A.H. Saman, M.Z.M., Kuan, Y.W., Hemdi, A.R., 2011. Integration model of fuzzy C-means clustering algorithm and TOPSIS method for customer lifetime value assessment, *Proc. IEEE Int. Conf. on Industrial Engineering and Engineering Management.*, pp. 1633-1637.
- Alborzi, M., 2008. Augmenting system dynamics with genetic algorithm and TOPSIS multivariate ranking module for multi-criteria optimization, *Proc. of Int. Conf. of the System Dynamics Society.*
- Hossen, J., Rahman, A., Sayeed, S., Samsuddin, K., Rokhani, F., 2011. A modified hybrid fuzzy clustering algorithm for data partitions, *Australian Journal of Basic and Applied Science*, 5(8): 674-681.
- Priyono, A., 2005. Generation of fuzzy rules with subtractive clustering, *Jurnal Teknologi*, 43(D): 143-153.
- Hung, T.-W., Fang, S.C., Nuttle, H.L.W., 1999. A clustering-based approach to fuzzy system identification, *Proc. Eighth Int. Fuzzy System Association World Congress.*, pp. 415-419.
- Yang, X.Y., Garibaldi, J.M., Bird, B., George, M.W., 2007. A novel fuzzy clustering algorithm for the analysis of axillaries lymph node tissue sections, *Journal Applied Intelligence*, 27(3): 237-248.
- Liao, T.W., Celmins, A.K., Hamell, R.J. 2003. A fuzzy c-means variant for the generation of fuzzy term sets, *Fuzzy Set and Systems*, vol. 135, pp. 241-257.
- Cintra, M.E., Camargo, H.A., Monard, M.C., 2008. A study on technique for the automatic generation of membership functions for pattern recognition., *Proc. III Congresso da Academia Trinacional de Ciencias.*, pp. 1-10.

